

1 A Novel Complotype Combination Associates with Age-
2 Related Macular Degeneration and High Complement
3 Activation Levels *in vivo*

4 Constantin C. Paun^{1,2}, Yara T.E Lechanteur¹, Joannes M. M. Groenewoud⁴, Lebriz Ersoy³, Tina
5 Schick³, Mohamed R. Daha⁵, Sascha Fauser³, Carel B. Hoyng¹, Anneke I. den Hollander^{1,2,6}, Eiko K.
6 de Jong*^{1,2}

7 ¹Department of Ophthalmology, Radboud University Medical Center, Philips van Leydenlaan 15,
8 6525 EX, Nijmegen, The Netherlands

9 ²Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Geert
10 Grooteplein 28, 6525 GA Nijmegen, The Netherlands

11 ³Department of Ophthalmology, University Hospital of Cologne, Kerpener Str. 62, 50924, Cologne,
12 Germany

13 ⁴Department for Health and Evidence, Radboud University Medical Center, Geert Grooteplein Noord
14 21, 6525 EZ, Nijmegen, The Netherlands

15 ⁵Department of Nephrology, Leiden University Medical Center, Albinusdreef 2
16 2333 ZA, Leiden, The Netherlands

17 ⁶Department of Human Genetics, Radboud University Medical Center, Geert Grooteplein 10, 6525
18 HB, Nijmegen, The Netherlands

19

20 Corresponding author: Eiko K de Jong,
21 Radboud University Medical Center, Department of Ophthalmology, internal zip: 409
22 Philips van Leydenlaan 15, 6525 EX, Nijmegen
23 The Netherlands
24 Eiko.deJong@radboudumc.nl

1 **Supplementary Table 1. Association between AMD and SNP genotypes in the *CFH*, *CFB*
2 and *C3* genes**

SNP	cDNA change	Protein change	Genotype	N		P-value	OR	95% C.I. for EXP(B)	
				AMD	Control			Lower	Upper
CFH rs800292	c.134G>A	p.V62I	GG	1101	801	7.4*10 ⁻¹¹			
			GA	445	539	0.015	0.63	0.435	0.913
			AA*	62	77	2.4*10 ⁻¹¹	0.564	0.477	0.667
CFB rs4151667	c.26T>A	p.L9H	TT*	1511	1291	0.02			
			TA	99	127	0.904	0.828	0.038	17.888
			AA	1	1	0.005	0.654	0.486	0.881
CFB rs641153	c.95G>A	p.R32Q	GG*	1436	1195	3*10 ⁻⁴			
			GA	171	217	0.387	0.545	0.138	2.152
			AA	4	7	7.9*10 ⁻⁵	0.623	0.492	0.788
C3 rs2230199	c.304G>C	p.R102G	CC*	911	901	2.4*10 ⁻⁶			
			CG	570	471	0.046	1.183	1.003	1.396
			GG	117	47	7.8*10 ⁻⁷	2.59	1.775	3.777

3
4 Analyses were performed by logistic regression analysis. The genotypes marked with * are
5 the ancestral variants. Variables entered in the model: *CFH* rs800292, *CFB* rs4151667, *CFB*
6 rs641153, *C3* rs2230199, age and gender. Bonferroni corrected threshold for statistical
7 significance is p<0.004.

8

1 **Supplementary Table 2. Genotype combination frequency for the novel complotype**

CFB (rs4151667) - CFB (rs641153) - CFH (rs800292)	Control		AMD		Total
	n	%	n	%	
AA - GG - GG	1	0.1	1	0.1	2
TA - GA - AA	0	0.0	2	0.1	2
TT - AA - GA	1	0.1	1	0.1	2
TA - GA - GA	5	0.4	1	0.1	6
TT - AA - GG	6	0.4	3	0.2	9
TA - GG - AA	7	0.5	3	0.2	10
TA - GA - GG	10	0.7	5	0.3	15
TT - GA - AA	11	0.8	9	0.6	20
TA - GG - GA	47	3.4	23	1.4	70
TA - GG - GG	55	3.9	65	4.1	120
TT - GG - AA	59	4.2	48	3.0	107
TT - GA - GA	74	5.3	47	2.9	121
TT - GA - GG	112	8.0	106	6.6	218
TT - GG - GA	406	29.0	370	23.1	776
TT - GG - GG	607	43.3	916	57.3	1523
Total	1401	100	1600	100	3001

2

3

4

1 **Supplementary Table 3. Differences in mean complement activation levels between**
 2 **genotype combinations**

<i>CFB</i> (rs4151667) - <i>CFB</i> (rs641153) - <i>CFH</i> (rs800292)	Mean Difference (I-J)	Std. Error	P-value
TA - GG - GA	TA - GG - GG	-0.024	0.026
	TT - GA - GA	-0.041	0.027
	TT - GA - GG	-0.085*	0.024
	TT - GG - AA	-0.06	0.027
	TT - GG - GA	-0.079*	0.022
	TT - GG - GG	-0.100*	0.021
TA - GG - GG	TA - GG - GA	0.024	0.026
	TT - GA - GA	-0.017	0.024
	TT - GA - GG	-0.061	0.021
	TT - GG - AA	-0.036	0.023
	TT - GG - GA	-0.055*	0.018
	TT - GG - GG	-0.076*	0.017
TT - GA - GA	TA - GG - GA	0.041	0.027
	TA - GG - GG	0.017	0.024
	TT - GA - GG	-0.044	0.021
	TT - GG - AA	-0.019	0.024
	TT - GG - GA	-0.038	0.018
	TT - GG - GG	-0.059*	0.018
TT - GA - GG	TA - GG - GA	0.085*	0.024
	TA - GG - GG	0.061	0.021
	TT - GA - GA	0.044	0.021
	TT - GG - AA	0.025	0.021
	TT - GG - GA	0.006	0.014
	TT - GG - GG	-0.015	0.014
TT - GG - AA	TA - GG - GA	0.06	0.027
	TA - GG - GG	0.036	0.023
	TT - GA - GA	0.019	0.024
	TT - GA - GG	-0.025	0.021
	TT - GG - GA	-0.019	0.018
	TT - GG - GG	-0.04	0.018
TT - GG - GA	TA - GG - GA	0.079*	0.022
	TA - GG - GG	0.055*	0.018
	TT - GA - GA	0.038	0.018
	TT - GA - GG	-0.006	0.014
	TT - GG - AA	0.019	0.018

	TT - GG - GG	-0.021	0.008	0.215
TT - GG - GG	TA - GG - GA	0.100*	0.021	6.1×10^{-5}
	TA - GG - GG	0.076*	0.017	1.9×10^{-4}
	TT - GA - GA	0.059*	0.018	0.0189
	TT - GA - GG	0.015	0.014	1
	TT - GG - AA	0.04	0.018	0.464
	TT - GG - GA	0.021	0.008	0.215

1 *The mean difference is significant at the 0.05 level. All p-values were adjusted for multiple
 2 comparisons: Bonferroni. The general linear model was corrected for age, gender, BMI and
 3 disease status.

4

5

6 **Supplementary Table 4. Amino acid conservation for *CFH* (rs800292, p.V62I) - *CFB*
 7 (rs4151667, p.L9H) - *CFB* (rs641153, p.R32Q) - *C3* (rs2230199, p.R102G)**

Species	CFH p.V62I	CFB p.L9H	CFB p.R32Q	C3 p.R102G
Human	V	L	R	R
Chimp	I	L	Q	R
Mouse	I	L	R	G
Dog	I	L	A	G
Cat		L	G	G
Cow		L	G	G

8

9

10 **Statistical Syntax used for the models built in SPSS and the R script used to run the
 11 Random forest analyses:**

12 SPSS syntax for the statistical models:

13 UNIANOVA Log_C3d_C3 BY Gender Disease_status Complotype_SNP2_SNP3_SNP4
 14 WITH Age_Blooddate Q14_BMI

15 /METHOD=SSTYPE(3)

16 /INTERCEPT=INCLUDE

17 /EMMEANS=TABLES(Complotype_SNP2_SNP3_SNP4) WITH(Age_Blooddate=MEAN
 18 Q14_BMI=MEAN) COMPARE

19 ADJ(BONFERRONI)

```

1 /PRINT=ETASQ PARAMETER
2 /CRITERIA=ALPHA(.05)
3 /DESIGN=Gender Disease_status Complotype_SNP2_SNP3_SNP4 Age_Blooddate
4 Q14_BMI.
5
6 LOGISTIC REGRESSION VARIABLES Disease_status
7 /METHOD=ENTER Complotype_SNP2_SNP3_SNP4 Age_Blooddate Gender
8 /CONTRAST (Complotype_SNP2_SNP3_SNP4)=Indicator
9 /CLASSPLOT
10 /PRINT=CI(95)
11 /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
12
13 The R script used for the random forest analyses
14 library(randomForest)
15 setwd("")
16 data <- read.table("file.txt", header=T)
17 View(data)
18 Combi3_4_SNPs <- data[, (2:11)]
19 attach(Combi3_4_SNPs)
20 set.seed(4)
21 complotype.rf <- randomForest(Log_C3d_C3 ~ ., data=Combi3_4_SNPs, mtry=2,
22 importance=TRUE)
23 print(complotype.rf)
24 round(importance(complotype.rf), 2)
25
26 library(randomForest)
27 setwd("")
```

```
1 data <- read.table("file2.txt", header=T)
2 View(data)
3 selected_columns <- data[, (2:11)]
4 names(selected_columns)
5 attach(selected_columns)
6 sink("results_randomForest_on_disease_status.txt")
7 set.seed(4)
8 complotype_on_AMD.rf <- randomForest(Disease_status ~ ., data=selected_columns,
9 importance=TRUE, proximity=TRUE)
10 print(complotype_on_AMD.rf)
11 round(importance(complotype_on_AMD.rf), 2)
```